Visual causality: investigating graph layouts for understanding causal processes

Dong-Bach Vo[1], Kristina Lazarova[1], Helen C. Purchase[1], and Mark McCann[2]

[1] School of Computing Science, University of Glasgow, Glasgow G128RZ, UK
[2] SMRC/CSO Social and Public Health Sciences Unit, University of Glasgow G23AX, UK
{firstname,lastname}@glasgow.ac.uk

Abstract. Causal diagrams provide a graphical formalism indicating how statistical models can be used to study causal processes. Despite the extensive research on the efficacy of aesthetic graphic layouts, the causal inference domain has not benefited from the results of this research. In this paper, we investigate the performance of graph visualisations for supporting users' understanding of causal graphs. Two studies were conducted to compare graph visualisations for understanding causation and identifying confounding variables in a causal graph. The first study results suggest that while adjacency matrix layouts are better for understanding direct causation, node-link diagrams are better for understanding mediated causation along causal paths. The second study revealed that node-link layouts, and in particular layouts created by a radial algorithm, are more effective for identifying confounder and collider variables.

Keywords: Causal inference · Causal graph · Graph layout

1 Introduction

Causal inference, used in areas as diverse as employment discrimination and biochemical reactions, is the study of whether a putative cause is responsible for an effect [7, 10]. A causal system can be expressed as a set of graphical objects: nodes, representing variables, with possible causal relationships from one to another represented by directed edges [24].

Causal diagrams provide specific graphical structures that facilitate the identification of specific causal model properties (see Fig. 1). In a causal graph, variables are represented with nodes, and statistical dependance, (i.e. causal relationships) between two variables with edges. A causal path is defined by an exposure, an outcome, and the set of all nodes and directed edges that connect the exposure to the outcome. Fig. 2 shows different causal paths from the exposure node A to the outcome node D. If a node on a causal path is caused by two other nodes on that same path, it is known within the social science community as a collider; the effect of this is that the statistical dependence between the two other nodes may be weakened. If a node on a causal path influences multiple

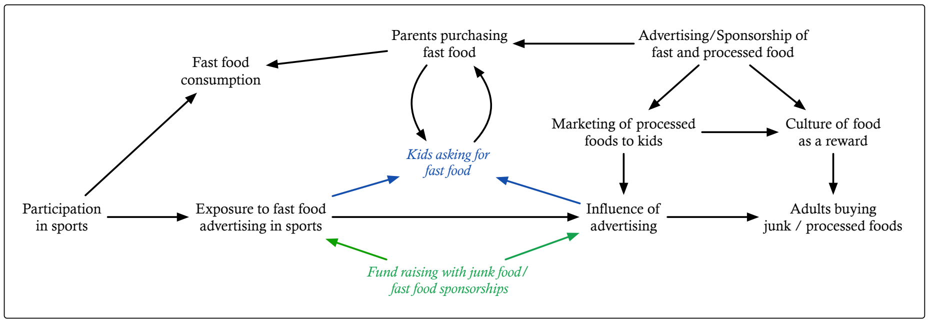

Fig. 1. A causal path from the Community Based Systems Diagram of Obesity Causes created by health and well being experts [1]. In the causal path between the nodes *Participation in sports* and *Adults buying junk/processed foods*, *Kids asking for food* is a collider and *Fund raising with junk food/fast food sponsorships* is a confounder.

other nodes on the same path, it creates a confounding bias: thus "back door paths", with such nodes are called confounders. These graphical structures give information on the influence of an external intervention on an outcome: in the first case, influencing A will lead to a corresponding change in D, whereas in the latter two cases, changing A may not cause a change in D. Identifying such graphical structures on small graphs is straightforward, however causal models and their graphical representation can be sophisticated and challenging to work with [13]. Fig. 1 shows an example of both a confounder and a collider.

Despite the prior extensive research on the relative usefulness of different graph layouts for a variety of tasks, the causal inference domain has not benefited from graph layout research. This avenue of research has the potential to have a significant impact on the way in which causal graphs are used in applied research and decision-making, for example in the formulation of health policy.

In this paper, we investigate how different graph visualisations can support causal reasoning. To the best of our knowledge, no other research has investigated this. In the first study, we investigate which layouts are most appropriate for

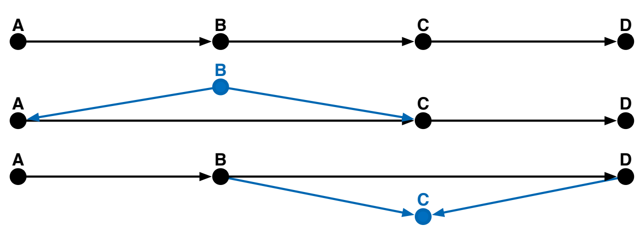

Fig. 2. Causal paths between A and D. Top: directed causal path. Center: causal backdoor path, B is a confounder on the path. Bottom: causal blocked path, C is a collider on the path.

studying causal paths and understanding causal relationships. The results show that adjacency matrix layouts yielded better performance for reasoning on direct causation and node-link layouts for reasoning on mediated causation. The second study investigated which node-link layout methods facilitate the identification of particular causal visual structures in graphs. Participants performed the best with radial layouts.

2 Related Work

This research aims to improve the visual approaches used in applied causal inference domains. It builds upon previous research in statistical causal inference using graphs and on research on visualising relationships in data.

2.1 Visualising Causal Inference

Causal graphs are networks that represent causation or the influence between properties of a domain. For example, the obesity system map represents influences such as education, stress or purchasing power over obesity [1,13]. Causation can be modelled quantitatively (the relationships between the entities are formalised in terms of conditional probability distributions derived from empirical data) or qualitatively (based on personal or expert opinion) [18].

Causal graphs formalize one's understanding of causal influences [24]. In population health, they have supported researchers to understand the associations between social policy, family characteristics, genetics, and foetal alcohol spectrum disorder [23]. While numerical statistical models can support causal inference, graphical approaches to causal problems have had a profound influence on the ways in which statistical models have been (and should correctly be) constructed [30], as well as providing a more engaging method of presenting evidence and eliciting opinions around causal questions with non-statistical audiences. Sophisticated interactive visualization applications exist to support causal inference using quantitative causal models represented as graphs [7, 27, 30, 32], e.g. *Tetrad* [27], *Dagitty* [30], *Visual Causal Analyst* (VCA) [32].

2.2 Representing a graph

The pioneering research in the graphical representation of causes was Wright's method in the field of animal genetics [34], formalising the influence of plausible causes on variables in a system combining mathematical and graphical modelling. The graphical model provides a causal overview as a directed acyclic graph where nodes represent causal variables and directed edges the causal relationships between variables. While later research contributed towards better and mathematically proven and graphical methodologies to measure causality [24], no empirical study has been conducted to evaluate the understandability of such graphical representations.

In contrast, the layout of directed graphs has been studied extensively in the graph drawing and information visualization research community [2, 14, 20], and several studies have found that the way in which a graph is laid out plays an important role in revealing the underlying meaning and structure of graphs [2]. For example, Purchase *et al.*'s study on the influence of aesthetic graphic layout criteria such as edge bending, edge crossings, or edge angle between nodes on graph readability, showed that edge crossings affect the graph reading most [25]. Another study looking at eye movements when reading graphs revealed that edge length may also affect performance [2].

The semantic domain of graphs should also be considered when designing or selecting layout [22, 25]. For instance, McGrath *et al.* found that participants perceived differently the 'prominence' and the 'bridging' properties of a social network depending on the position of the nodes in undirected graphs [22], concluding that, given a specific domain, the best representation may depend on the type and the valence of the information one wants to convey. Causal diagrams are semantically rich as they can communicate probabilistic independence or show confounding biases, and no research to date has investigated the best graph layout to support the understanding of causal diagrams.

As an alternative directed graph representation, adjacency matrices show relationships between nodes in a binary matrix, with target and source nodes of each edge indicated in the matrix cells. (Fig. 3).

Interaction with adjacency matrices has been found to be worse than with node-link diagrams [11, 12]. Ghoniem *et al.* showed that participants performed better in several topographic graph reading tasks using matrices [12]. The task of finding paths between two nodes was better using node-link diagrams, though this performance decreased as the size of the graphs increased. Keller *et al.* generalised Ghoniem et al's finding [19], suggesting that the suitability of the representation may depend on the task performed and its semantic nature [19].

Information visualisation systems can combine adjacency matrix with node-link layouts; for example, *MatrixExplorer* offers a way to switch from matrix to node-link to take advantage of both representations [15]. Both representations can be used to depict different types of relationships. *NodeTrix* visualises social networks and performs very specific tasks relating to social sciences: the matrix layout represents intra-community relationships while node-links layout is used to depict inter-community relationships [16].

2.3 Comparing layouts for causal inference

The most efficient layouts for causal inference may depend upon nature of the causal reasoning tasks. Since node-link diagrams are the most common graphical representation for causal inference, layouts implementing the best for graph reading, such as minimisation of edges crossing or orthogonality [25], could improve causal reasoning task performance. Adjacency matrices have been shown to outperform node-link diagrams for many abstract related tasks but such studies have not been conducted on a semantically-rich directed graphs like those used in causal reasoning [12]. We report on two studies which aim to compare

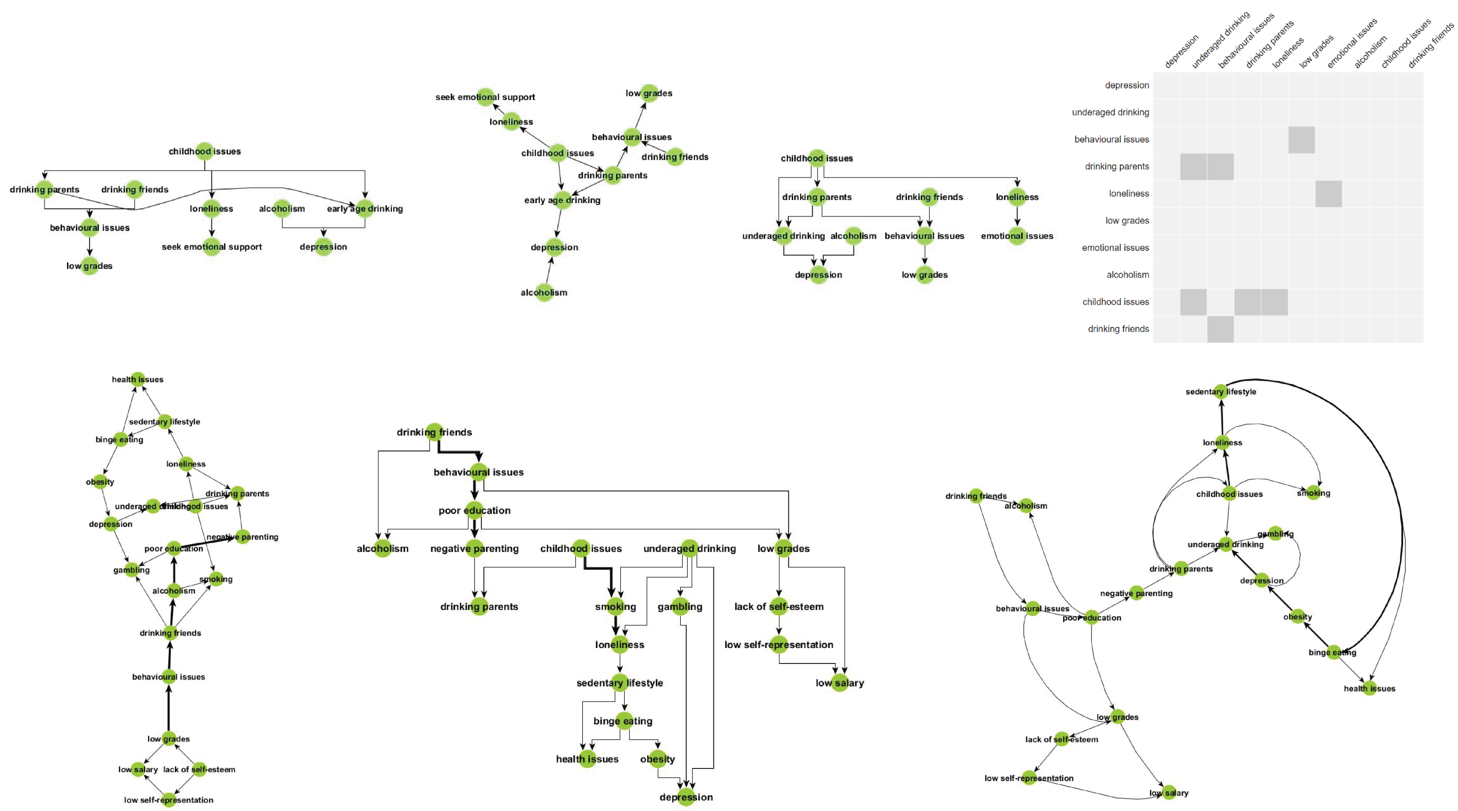

Fig. 3. Types of layout in the study. The top line shows some layouts used to investigate causation intelligibility. From left to right, the graph layouts are: parallel-series (PL), spring (SL), hierarchical (HL) and matrix out-degrees descending order (MODL). The bottom line shows some layouts to investigate the identification of causal structures. From left to right, the graph layouts are: spring (SL), hierachical (HL) and radial (RL).

graph layouts for causal inference. The first study investigates the best visualisation method for understanding causal paths in causal graphs; the second study compares the use of different graph layouts in identifying causal structures.

3 Investigating causation intelligibility

We investigated task performance when participants explore a causal graph when answering questions about its causal path relationships, looking at three common tasks related to causal inference: understanding direct causation, understanding mediated causation (i.e. indirect causation), and identifying causal structures.

Several node-link layouts have been proposed in the graph drawing community: we selected those we believed would improve participants' performance [8] (Fig. 3). The *hierarchical* layout emphasises structures in graphs by following regular patterns that can be easily followed by users' eyes [29]: including drawing direct connected nodes close to each other, limiting the number of edge crossings, and an orthogonal layout. *ReactionFlow*, a tool to support causal inference in biology, inspired the choice of the *parallel-series* directed graphs layout, popular for visualising flows in data [7]. This layout combines several graphs by merging the common roots into a single root when possible with all the paths parallel to each other, minimising edge crossings and bending, and following an orthogonal form shown to be effective for understanding abstract graphs [2,4,25]. In *spring* layouts, physical repulsive forces result in nodes with weak ties being pulled away from the others. Since as confounders and colliders on the graph are attached to a causal path by at least two edges going to the same direction (Fig. 1), this type of layout might create highly visible clusters around such highly connected nodes.

Several reordering techniques for highlighting data of interest through visual patterns in adjacency matrices have been proposed [3,21]. *Alphabetic* layouts have been found to outperform node-link layouts for tasks related to reading undirected graphs larger than 20 nodes [12], and can improve graph reading performance especially for users without prior knowledge of a domain [19]. Two other matrix layouts were added: *out-degrees* and *in-degrees* descending arrangement, with the expectation that they could help identify colliders and confounders. The *out-degrees* (resp. *in-degrees*) descending arrangement sorts the number of the edges going out from (resp. going towards) each vertex in a descending order.

4 Investigating the effect of causal layouts

We designed 3 datasets each from one of these themes: drinking issues, examinations, and health related gym behaviour; for each, we created graphs of different sizes: 10, 20 and 30 nodes. Each graph contains several causal paths, where a causal path is a path through the graph from an exposure node (e.g. teenage drinking), through mediated nodes (e.g. alcohol dependency, depression, liver failure) to an outcome node (e.g. death from alcoholism). All causal paths in the graphs included 8 nodes and 7 edges. We had 6 presentation conditions, 3 node-link drawings (*spring*, *hierarchical*, *parallel*) and 3 matrix presentations (*alphabetic*, *in-degree*, *out-degree*).

Two types of question were used: direct causation on a path (e.g. "What factor is causing factor X?"), and mediated causation along longer paths ("Is this causal path correct?"). It has been shown that following mediated paths in applied causal contexts is not trivial [5]. One question of each type was associated to each possible graph (3 *sizes* $\times$ 6 *presentations*) that being 36 unique tasks.

We anticipated that node-link layouts would result in better performance for understanding causation (H1) since matrix layouts do not perform well for following paths in abstract undirected graphs of over 30 nodes [12]. In particular, *hierarchical* layout would be the best layout (H2) as it has been proven to be successful for abstract graphs [25].

4.1 Experimental design

The yEd Graph Editor was used to create the graphs and the layouts with respect to the three chosen node-link layout algorithms [35]: *hierarchical*, *parallel-series*, and *spring* (yEd's organic force-directed layout). The adjacency matrices were arranged with *alphabetic*, *in-degrees* and *out-degrees* descending orders (Fig. 3).

4.2 Procedure

The experiment was conducted using a custom-built experimental software on a laptop computer, in the presence of the experimenter. The training materials (written documentation and video) presented to participants had been piloted with several people in advance to ensure that they adequately explained the task and did not include obvious biases. These materials used a graph of only five nodes and four edges to explain the concept of causality between variables and how it is depicted in both node-link and graphical form. They were then invited to ask for any clarification.

For each trial, the stimulus consisted in displaying a layout among the 18 available and one of the two associated questions, together with multiple choice options for answering the question. For each question, 4 potential answers were suggested to the participants. Participants were told that the correct answer could always be found in the graph and that they should not need to resort to guessing. The plausible answers were presented through a radio-button list to guarantee a unique answer from the participants. The trial was over when the participant selected an answer by clicking on a radio button.

The experiment started with 6 training questions, which were discarded from the dataset, to help them familiarise themselves with the system and the task, and to mitigate any learning effect. The experimental phase comprised 36 trials with stimuli ordered by a partial Latin square design to avoid any presentation order effect. Since we wanted to emulate a reader's process of attempting to understand a causal graph as a whole, participants had to visually scan the drawing to identify the nodes of interest before responding. At the end participants were given a questionnaire to assess subjective preferences. Each evaluation session lasted for about 60 minutes.

The independent variables were layout type, graph size and question type. The dependent variables were response time and answer accuracy. The response time was measured between the stimulus appearance and the validation of an answer, including the duration of the cognitive process of understanding the causal question and the localisation of the nodes of interest. Thirty volunteers took part in the study. Participants were between 20 and 29 years old ($Mdn = 22.57$, $SD = 2.14$) and 6 were male. They were all undergraduate students with no prior experience working with graphs.

4.3 Results

We collected 1080 trials (30x36), with success rate of 98.80%. We removed the data from 12 outlying trials, when the distance of the sample from the mean response time was three times greater than a standard deviation (i.e. greater than 86s). To accommodate any non-parametric nature of data distribution, an aligned rank transform (ART) was performed before further analysis [17, 33].

Table 1. Median response time in seconds by graph layout and size for direct and mediated causal inference. S denotes small graphs, M medium graphs and L large graphs. HL is hierarchical, SL is spring and PL is parallel layout. MAL is matrix alphabetic, MIDL is matrix in-degree and MODL matrix out-degree ordering layout.

Size		Direct causation						Mediated causation					
		HL	SL	PL	MAL	MIDL	MODL	HL	SL	PL	MAL	MIDL	MODL
S	Median	8.81	11.36	13.19	10.57	9.97	8.42	26.17	16.54	20.35	13.64	21.96	26.22
	IQR	4.75	3.20	4.95	5.20	4.88	4.01	7.83	6.99	8.6	11.56	14.95	13.33
M	Median	13.54	12.85	12.07	12.46	13.88	11.06	12.49	37.19	22.72	19.01	33.93	23.92
	IQR	8.53	5.72	6.54	10.11	5.18	4.60	6.21	12.93	11.42	10.15	24.65	17.91
L	Median	15.45	16.94	26.32	14.06	13.59	13.30	25.83	45.89	27.79	44.46	42.13	44.83
	IQR	12.39	8.13	8.73	6.7	4.59	7.23	16.84	26.70	37.37	16.46	18.10	15.13

The two questions asked (direct causation between two nodes, medicated causation along a path) are sufficiently different for separate analyses to be appropriate.

Direct causation. The median response time was the fastest for MODL for small, medium and large graphs. The slowest response times were found for PL with small and large, and for MIDL for medium graphs (Table 1 and Fig. 4).

A two-way ANOVA on the aligned rank transformed data revealed a significant main effect with layout ($F_{5,492} = 20.21$, $p < .0001$) and size ($F_{2,492} = 105.44$, $p < .0001$) factors, and an interaction effect for layout × size ($F_{10,492} = 9.91$, $p < .0001$). A *Post hoc* Tukey's HSD test found all the layouts significantly faster than PL (all $t(491) > 5.91$, $p < .0001$), and MODL significantly faster

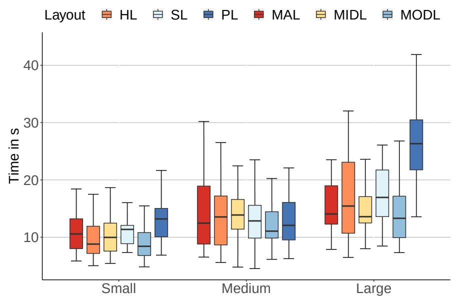

Fig. 4. Median response times for direct causation

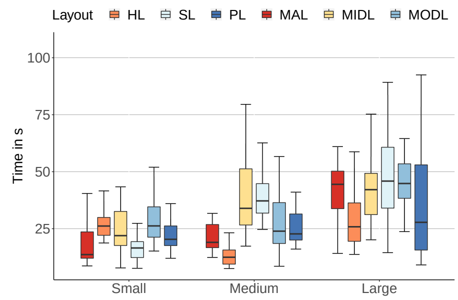

Fig. 5. Median response times for mediated causation

than SL ($t(491) = 3.30$, $p < .001$) and MAL ($t(491) = 3.05$, $p < .05$). Exploring small graphs was significantly faster than medium ($t(493) = 5.48$, $p < .001$) and large ($t(493) = 14.39$, $p < .001$) graphs; Exploring medium faster than large graphs ($t(493) = 8.92$, $p < .001$). *Post hoc* pairwise comparisons with Holm-Bonferroni correction revealed that all the interactions of all the layouts with PL were significant for small and large (all $\chi^2 > 33.58$, $p < .0001$), and for medium and large graphs (all $\chi^2 > 26.06$, $p < .0001$).

Mediated causation. The fastest response time was found for MAL with small graphs, for HL with medium and large graphs. It was the slowest for MODL with small graphs and for SL with medium and large graphs. The results are shown in Table 1 and in Fig. 5.

An ANOVA on the align rank transformed data showed a significant main effect for layout ($F_{5,482} = 19.07$, $p < .0001$) and size ($F_{2,482} = 82.13$, $p < .0001$) and a significant effect for interaction of both ($F_{10,482} = 12.64$, $p < .0001$). A Tukey's *Post hoc* pairwise comparison found MAL significantly faster than HL, MIDL, SL and MODL (all $t(482) > 3.37$, $p < .05$). HL was significantly faster than MIDL, SL, and MODL (all $t(482) > 7.02$, $p < .0001$) and PL faster than MIDL, SL and MODL (all $t(511) > 4.54$, $p < .001$). Exploring small graphs was significantly faster than medium ($t(482) = 4.69$, $p < .0001$) and large ($t(482) = 12.68$, $p < .0001$) graphs, and exploring medium faster than large ($t(481) = 8.08$, $p < .0001$). *Post hoc* pairwise comparison of factor interactions using Holm-Bonferroni correction showed significant differences between medium and large graphs for MAL with MIDL, SL and PL (all $\chi^2 > 12.80$, $p < .01$). It was also the case for MODL with MIDL, SL and PL (all $\chi^2 > 12.86$, $p < .01$). The differences between small and large graphs were significant for HL with MAL, MIDL, MODL and SL (all $\chi^2 > 12.22$, $p < .05$), for MAL with PL ($\chi^2 18.81$, $p < .001$), and for SL with MIDL and PL (both $\chi^2 > 9.70$, $p < .05$). There were significant differences between medium and large graphs for SL with HL, MAL, MODL and PL (all $\chi^2 > 16.54$, $p < .005$), for HL and MAL, MIDL and PL (all $\chi^2 > 18.59$, $p < .005$), and for MODL with MIDL ($\chi^2 11.53$, $p < .05$).

Participants completed a questionnaire investigating the relative easiness of working with node-link or matrix layouts on a scale from 0 to 5. A Mann-Whitney test indicated that the median score for node-link layout ($Mdn = 3$, $IQR = 2$) was significantly greater than the score for matrix layout ($Mdn = 2$, $IQR = 1.75$, $U = 234.5$, $p < 0.001$). 80% of the participants preferred to work with node-link diagrams over matrices.

4.4 Discussion

H1 was not supported: node-link diagrams were not the fastest. PL and SL exhibited the worst performance for understanding direct and mediated causation in large graphs. H2 was partially supported as HL gave the best performance for understanding mediated causation but only with medium and large graphs.

The results indicated an interesting trend when comparing layout performance with regard to graph sizes. Some layouts that were slower for small graphs were faster with medium or large graphs. MIDL became significantly faster than some node-link layouts with small and large graphs for understanding direct causation. However, the participants did not notice this performance boost with matrix layouts; subjective ratings show preference for node-link layouts.

Performance also differed for direct and mediated causal reasoning. MODL showed the best performance for direct causal reasoning whatever the graph size, outperforming all node-link layouts. Finding a direct cause was easier on matrix layouts than on node-link layouts, contradicting H2. While labels are scattered around the plane in node-link layouts, they are arranged following a single horizontal (columns) or vertical (rows) line in matrix layouts. Furthermore, the fast access to highly connected nodes supported participants in highlighting causal relationships, and arranging the row and the column headers with alphabetic order helps users to locate the target nodes and its causal predecessor or successor even faster. For small and large graphs, the *parallel-series* layout displayed the worst performance, showing that even if paths are explicitly drawn, not all the node-link layouts are appropriate for causal inference. This is an important finding as this layout is currently used in existing causal inference software [7].

For mediated causation reasoning, as graph sizes increased, the overall performance got worse, but the worsening in performance for each layout differed. HL was the fastest for medium and for large graphs, making it a good option for reasoning about mediated causation and partially validating H2. While it seems the alphabetic matrix layout was the most efficient for small graphs, node-link based layouts were faster for medium and large graphs. Because matrix layouts require users to perform saccades from row to column headers to follow paths, analysing long paths were more challenging. This was accentuated by the fact that, depending on the arrangement of the rows and the columns, two consecutive nodes' labels are unlikely to be located next to each other in the matrix headers. These results are in line with the findings for syntactic graphs [12, 28] and connectivity models [19].

The results suggest MIDL may be promising for causal reasoning in even larger graphs than in this study; further research is needed to confirm this.

5 Identifying causal structures

The first study focused on understanding which layout best supports following directed causal paths. However, the presence of colliders or confounders can affect causal interpretation, conflicting with intuition. Highlighting these causal processes effectively is crucial for evidence-informed decision making.

5.1 Experimental design

Two sets of questions were designed: investigating the direct identification of confounders and colliders relative to a path, and identifying these causal structures by exploring the entire graph.

After being presented with a highlighted pair of one node and one path in a graph, *direct identification* question asked whether the node was a collider or a confounder with respect to the path, or neither. For the *exploratory identification* question, a path of a graph was highlighted and participants were asked to enumerate all the colliders and confounders related to the path.

The hierarchical (HL) and spring (SL) layouts from the previous study were retained, but the parallel-series layout (PL) was discarded because of its weak performance in the first study. A radial tree layout (RL, also created by yEd, as noted in section 4.1) was added to the conditions as this layout is widely used to depict relationships among diverse entities [9]. An adjacency matrix layout was also included, with rows and columns ordered by the type of the nodes: the nodes at the start of the causal paths were used as first indices, while the nodes at the end of the causal paths were used as the last indices. This was so as to gather meaningful causal information in the centre of the layout as much as possible. Only two graph sizes were used (medium: 20 nodes, large: 40 nodes). Small graphs were discarded as we thought the task would be too easy given the results of the first study. For each domain (drinking issues, exams, and health related gym behaviour), 6 new graphs were generated. Each causal path in the graphs included 8 nodes and 7 edges.

5.2 Procedure

Participants were introduced to causal relationships and their representation with node-link and matrix layouts. Then, the experimenter explained to the participants the collider and confounder concepts and what these structures look like on node-link and matrix diagrams: a collider on a causal path is a node resulting of a common effect of two other nodes on this same path; a confounder on a causal path is a node that influences multiple other variables on this same path (Fig. 1). Before starting, the participants could practice with two node-link and two matrix layout examples to ensure they had correctly understood the concepts and the instructions.

For each trial, the stimulus consisted in displaying a random question and the associated layout. For a direct identification task, participants had to select whether the node was a "collider", "confounder", or "none". For an exploration

Table 2. Mean success rate in % and median reaction time in seconds for direct identification of graphical causal structures.

Size		Success Rate					Reaction Time			
		HL	SL	RL	ML		HL	SL	RL	ML
M	Mean	92.13	92.22	87.64	65.17	Median	19.52	21.76	21.78	31.76
	SD	0.27	0.28	0.33	0.48	IQR	12.10	18.63	16.44	17.02
L	Mean	85.39	93.33	93.26	81.93	Median	21.34	17.50	20.79	33.00
	SD	0.35	0.25	0.25	0.39	IQR	16.95	0.25	18.71	18.72

task, they had to select the only four collider or confounder nodes with respect to the highlighted path among a list of 10 candidates. A partial Latin square was used to avoid any ordering effect. Each evaluation session lasted for about 60 minutes. The apparatus from the previous study was used.

The independent variables of this study were layout type, graph size and question type. The dependent variables were the response time, which was measured between the stimulus appearance and the validation of an answer, and the answer accuracy. A total of 540 answers were collected. Thirty volunteers took part to the study. Participants were between 20 and 27 years old ($M = 22.57$, $SD = 2.14$). All of them were students and 14 were female, and undergraduate students with no prior knowledge of graphs.

5.3 Results

The data were split according to the type of question (direct identification or exploration).

Direct identification of causal structures. We discarded 12 samples from our data for the direct identification task and 24 samples for the graph exploration task because their distance to the mean response time was greater than three times the standard deviation (i.e. greater than $224s$).

The success rate varied between 65.17% (ML) and 92.22% (SL) for medium sized graphs and between 81.93% (ML) and 93.33% (SL) for large sized graphs (Table 2). A repeated-measures ANOVA on the regression model found a significant difference for layouts ($F_{3,703} = 31.43$, $p < .001$). *Post hoc* Tukey's pairwise comparisons showed that ML was significantly worse than all the other layouts (all $p < .001$). Median response times for the direct identification task shown in ranged from $19.52s$ (HL) to $31.76s$ (ML) for medium graphs and from $21.34s$ (HL) to $33.00s$ (ML) for large graphs (Table 2). A two-way ANOVA on the ART data revealed a significant effect of the layout ($F_{3,203} = 31.75$, $p < 0.001$). *Post hoc* Tukey's pairwise comparisons found ML significantly slower than all the other layouts ($p < 0.001$).

Exploratory identification of causal structures. We discarded all the matrix data, since participants' performance in this condition was extremely poor, and no meaningful comparisons could be made. The success rate for finding colliders was 87.78% for medium graphs and 87.64% for large graphs, and 90% for confounders in medium graphs and 86.04% in large graphs (Table 3). A repeated measures ANOVA on the regression model showed significant effect of the layout ($F_{2,527=6.06}$, $p < 0.05$) for finding colliders. *Post hoc* Tukey HSD pairwise comparisons found SL significantly better than HL ($p < .05$). A significant effect of the size ($F_{1,529} = 8.07$, $p < .01$) and the layout ($F = 18.95$, $p < .001$) was found for finding confounders. *Post hoc* Tukey HSD pairwise comparisons found SL and RL significantly better than HL (both $p < .01$) and medium significantly better than large graphs ($p < .01$). The median response time for finding all the colliders and the confounders was $63.40s$ ($IQR = 47.28$) for SL, $72.94s$ ($IQR = 39.08$) for HL, and $82.35s$ ($IQR = 44.04$) for RL in medium graphs. It reached $70.19s$ ($IQR = 32.40$) for RL, $84.40s$ ($IQR = 50.71$) for HL and $83.11s$ ($IQR = 42.60$) for SL in large graphs. A two-way ANOVA on the ART data revealed a significant effect of size ($F_{1,145} = 13.62$, $p < .001$) and of the interaction between both factors ($F_{2,145} = 13.92$, $p < .0001$). *Post hoc* Tukey's pairwise comparisons found exploring medium graphs significantly faster than large graphs ($p < 0.001$). *Post hoc* pairwise comparison using Holm-Bonferroni correction based on the interaction revealed that while RL was slower than HL ($\chi^2 = 12.30$, $p < .001$) and SL ($\chi^2 = 26.71$, $p < .0001$) for medium graphs, it became faster than both for large graphs.

5.4 Discussion

The results for the matrix layout were so poor in supporting participants' identification of collider or confounder structures in graphs that we omitted them from the data analysis for both questions. This is interesting in itself, because not only were MIDL and MODL found to be promising for reasoning causal graphs in our first study, but also previous research has advocated for the usage of matrix layouts for reading nodes' connectivity in graphs of 20 nodes and more [12]. One possible reason may be our participants' unfamiliarity with the matrix representation, or the fact that there is no obvious visual pattern that clearly highlights the existence of confounders or colliders in matrices.

When exploring the graph to find confounders and colliders, HL exhibited the worst performance, despite being one of the most praised layouts for its aesthetic characteristics [6]. While SL and RL manifested similar accuracy for finding causal structures in graphs, RL performed better as the number of nodes

Table 3. Mean success rate in % for exploratory identification of causal structures in graphs. SD values are indicated in parentheses.

Node Type	HL Medium	SL Medium	RL Medium	HL Large	SL Large	RL Large
Colliders	81.11 (0.39)	87.78 (0.33)	82.02 (0.39)	74.71 (0.44)	87.21 (0.33)	87.64 (0.33)
Confounders	81.11 (0.39)	90.00 (0.30)	87.64 (0.33)	60.92 (0.49)	86.04 (0.35)	83.14 (0.38)

increased, becoming faster than both SL and HL with large graphs. This is a compelling finding as previous research on syntactic graphs has advised the use of RL over orthogonal layouts. This suggests that RL could better support users with identifying confounding and colliding processes.

6 General discussion and future work

While HL was the most efficient for understanding causal paths, matrix layouts were promising. In this context, MAL can improve the understanding of causation, and in particular, localisation on the nodes of interest. Note that MIDL performance also increased with graph size. MIDL and MODL give fast access to highly connected nodes — the nodes likely to be of interest in the causal inference process. However, these matrix layouts did not support the identification of causal structures which are likely to be highly connected. This may have been caused by the lack of expertise of our participants in causal inference and information visualisation. Further research is needed to understand better the potential of such layouts with directed graphs for causal inference in applied settings and especially with expert users.

None of the matrix layouts presented here were suitable for identifying causal structures. However, since row and column permutations affect readability, more research is needed to identify further permutations that might highlight causal relationships and similarities [11].

For node-link diagrams, we find that the RL node-link layout was the most efficient layout for identifying causal structures, but following causal graphs and identifying relevant structures to identify colliders and confounders would require different layouts. Another research direction might be to investigate how hybrid methods or animation could support users for juxtaposing or switching from one layout to another [14, 31].

Finally, we only looked at a limited set of causal structures, and limited path lengths; we thus have no way of knowing how layout features will operate under more complex and diverse circumstances that are likely to arise in applied settings. This limitation makes further research based on increasingly complex causal structures all the more important.

7 Conclusion

This is the first empirical study of how visual aesthetics can influence how non-expert viewers interpret causal graphs. Our findings suggest that existing principles for general graph readability are insufficient to depict causal graphs effectively. First, causal graphs have structures with a specific interpretation that do not appear in graphs used in other domains. Second, the domain problem is a compound sequence of basic visual analytic tasks (e.g. search the plane, identify connections, infer direction of connections). It appears that different layouts are faster for each basic task, and that there are unexpected relationships between the compound tasks and features of the layout.

Our findings suggest that matrix layouts are the best layouts to investigate direct causal relationships, with matrix-out-degree the fastest, while node-link diagrams with hierarchical layout is the most promising for mediated causation. For identifying causal structures, radial was the most promising layout, with its performance increasing with the size of graphs. This suggests that causal inference could benefit from visualisation tools that provide multiple coordinated views [26], thus supporting users in a range of different tasks for understanding causation. Further investigation that considers cognitive and visual processes would help in explaining the results of our experiment, and better understanding of the principles of visual causal inference will assist in developing readable and informative causal graphs.

Note

Ethical clearance was given by the Ethics Committee of the College of Science and Engineering at the University of Glasgow (ref: 300150001). Study materials are available at `http://www.dcs.gla.uk/~hcp/Diagrams2020`.

References

1. Allender, S., et al.: A Community Based Systems Diagram of Obesity Causes. PLOS ONE **10**(7), e0129683 (2015)
2. Bae, J., et al.: Developing and evaluating quilts for the depiction of large layered graphs. IEEE TVCG **17**(12), 2268–2275 (2011)
3. Behrisch, M., et al.: Matrix Reordering Methods for Table and Network Visualization. Computer Graphics Forum **35**(3), 693–716 (2016)
4. Bennett, C., et al.: The aesthetics of graph visualization. Proc. of the 2007 Computational Aesthetics in Graphics, Visualization, and Imaging pp. 57–64 (2007)
5. Braveman, P., Gottlieb, L.: The Social Determinants of Health: It's Time to Consider the Causes of the Causes. Public Health Reports **129**(2), 19–31 (2014)
6. Burch, M., et al.: Evaluation of Traditional, Orthogonal, and Radial Tree Diagrams by an Eye Tracking Study. IEEE TVCG **17**(12), 2440–2448 (2011)
7. Dang, T., et al.: ReactionFlow: an interactive visualization tool for causality analysis in biological pathways. BMC Proceedings **9**, S6 (2015)
8. Di Battista, G.: Graph Drawing: Algorithms for the Visualization of Graphs. An Alan R. Apt Book, Prentice Hall (1999)
9. Draper, G., et al.: A Survey of Radial Methods for Information Visualization. IEEE Transactions on Visualization and Computer Graphics **15**(5), 759–776 (2009)
10. Elmqvist, N., Tsigas, P.: Causality visualization using animated growing polygons. In: Proc. of IEEE Info. Vis. 2003. vol. 2003, pp. 189–196. IEEE (2003)
11. Garaigordobil, M., et al.: Childhood depression: Relation to adaptive, clinical and predictor variables. Frontiers in Psychology **8**(MAY), 1–9 (2017). https://doi.org/10.3389/fpsyg.2017.00821
12. Ghoniem, M., et al.: A comparison of the readability of graphs using node-link and matrix-based representations. Proc. of IEEE Info. Vis. 2004 pp. 17–24 (2004)
13. Government, U.: Reducing obesity: obesity system map, Tackling obesities: future choices building the obesity system map. `"https://www.gov.uk/government/publications/reducing-obesity-obesity-system-map"` (2007)

14. Heer, J., Robertson, G.: Animated Transitions in Statistical Data Graphics. IEEE TVCG **13**(6), 1240–1247 (2007)
15. Henry, N., Fekete, J.D.: MatrixExplorer: A Dual-representation system to explore social networks. IEEE TVCG **12**(5), 677–684 (2006)
16. Henry, N., et al.: NodeTrix: a Hybrid Visualization of Social Networks. IEEE Transactions on Visualization and Computer Graphics **13**(6), 1302–1309 (2007)
17. Higgins, J.J., Tashtoush, S.: an aligned rank transform test for interaction
18. Keatley, D.A., et al.: Lay understanding of the causes of binge drinking in the United Kingdom and Australia: a network diagram approach. Health Education Research **32**(1), cyw056 (2017)
19. Keller, R., et al.: Matrices or Node-Link Diagrams: Which Visual Representation is Better for Visualising Connectivity Models? Information Visualization **5**(1), 62–76 (2006)
20. von Landesberger, T., et al.: Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. Computer Graphics Forum **30**(6), 1719–1749 (2011)
21. Liiv, I.: Seriation and matrix reordering methods: An historical overview. Statistical Analysis and Data Mining **8**(5), 70–91 (2010)
22. McGrath, C., et al.: The effect of spatial arrangement on judgments and errors in interpreting graphs. Social Networks **19**(3), 223–242 (1997)
23. McQuire, C., et al.: The causal web of foetal alcohol spectrum disorders: a review and causal diagram. European Child & Adolescent Psychiatry (2019)
24. PEARL, J.: Causal diagrams for empirical research. Biometrika **82**(4), 669–688 (dec 1995)
25. Purchase, H.C., et al.: Empirical evaluation of aesthetics-based graph layout. Empirical Software Engineering **7**(3), 233–255 (2002)
26. Roberts, J.C.: State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In: Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007). pp. 61–71. IEEE (2007)
27. Scheines, R., et al.: The TETRAD Project: Constraint Based Aids to Causal Model Specification. Multivariate Behavioral Research **33**(1), 65–117 (1998)
28. Shen, Z., Ma, K.L.: Path Visualization for Adjacency Matrices. In: Museth, K., et al. (eds.) Eurographics/ IEEE-VGTC Symposium on Visualization. The Eurographics Association (2007)
29. Sugiyama, K., Misue, K.: Visualization of structural information: automatic drawing of compound digraphs. IEEE Transactions on Systems, Man, and Cybernetics **21**(4), 876–892 (1991)
30. Textor, J., et al.: DAGitty: a graphical tool for analyzing causal diagrams. Epidemiology (Cambridge, Mass.) **22**(5), 745 (2011)
31. Vehlow, C., Beck, F., Weiskopf, D.: The State of the Art in Visualizing Group Structures in Graphs. In: Borgo, R., et al. (eds.) Eurographics Conference on Visualization (EuroVis) - STARs. The Eurographics Association (2015)
32. Wang, J., Mueller, K.: The Visual Causality Analyst: An Interactive Interface for Causal Reasoning. IEEE TVCG **22**(1), 230–239 (2016)
33. Wobbrock, J.O., et al.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proc. of CHI '11. p. 143. ACM Press, New York, New York, USA (2011)
34. Wright, S.: Correlation and causation. Journal of agricultural research **20**(7), 557–585 (1921)
35. yWorks: yEd Graph Editor. `"http://www.yworks.com/products/yed"`